

06/19/00
JCS59 U.S. PTO

06/21/00

A

Please type a plus sign (+) inside this box → ☐

PTO/SB/05 (12/97)
Approved for use through 09/30/00. OMB 0651-0032
Patent and Trademark Office: U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it displays a valid OMB control number.

UTILITY PATENT APPLICATION TRANSMITTAL <small>(Only for new nonprovisional applications under 37 CFR 1.53(b))</small>	Attorney Docket No.	K35A0574	Total Pages	
	First Named Inventor or Application Identifier			
	BRUCE E. ALDRIDGE			
	Express Mail Label No.	EI266879917US		

APPLICATION ELEMENTS <i>See MPEP chapter 600 concerning utility patent application contents.</i>	ADDRESS TO: Assistant Commissioner for Patents Box Patent Application Washington, DC 20231
<p>1. <input checked="" type="checkbox"/> Fee Transmittal Form <i>(Submit an original, and a duplicate for fee processing)</i></p> <p>2. <input checked="" type="checkbox"/> Specification [Total Pages 28] <i>(preferred arrangement set forth below)</i></p> <ul style="list-style-type: none">- Descriptive title of the Invention- Cross References to Related Applications- Statement Regarding Fed sponsored R & D- Reference to Microfiche Appendix- Background of the Invention- Brief Summary of the Invention- Brief Description of the Drawings <i>(if filed)</i>- Detailed Description- Claim(s)- Abstract of the Disclosure <p>3. <input checked="" type="checkbox"/> Drawing(s) <i>(35 USC 113)</i> [Total Sheets 9] <input checked="" type="checkbox"/> Formal <input type="checkbox"/> Informal</p> <p>4. <input type="checkbox"/> Oath or Declaration [Total Pages 3]</p> <p>a. <input checked="" type="checkbox"/> Newly executed (original or copy)</p> <p>b. <input type="checkbox"/> Copy from a prior application (37 CFR 1.63(d)) <i>(for continuation/divisional with Box 17 completed)</i> <i>[Note Box 5 below]</i></p> <p>i. <input type="checkbox"/> DELETION OF INVENTOR(S) Signed statement attached deleting inventor(s) named in the prior application, see 37 CFR 1.63(d)(2) and 1.33(b).</p> <p>5. <input type="checkbox"/> Incorporation By Reference <i>(useable if Box 4b is checked)</i> The entire disclosure of the prior application, from which a copy of the oath or declaration is supplied under Box 4b, is considered as being part of the disclosure of the accompanying application and is hereby incorporated by reference therein.</p>	<p>6. <input type="checkbox"/> Microfiche Computer Program <i>(Appendix)</i></p> <p>7. Nucleotide and/or Amino Acid Sequence Submission <i>(if applicable, all necessary)</i></p> <p>a. <input type="checkbox"/> Computer Readable Copy</p> <p>b. <input type="checkbox"/> Paper Copy (identical to computer copy)</p> <p>c. <input type="checkbox"/> Statement verifying identity of above copies</p>
ACCOMPANYING APPLICATION PARTS	
<p>8. <input checked="" type="checkbox"/> Assignment Papers (cover sheet & document(s))</p> <p>9. <input type="checkbox"/> 37 CFR 3.73(b) Statement <input checked="" type="checkbox"/> Power of Attorney <i>(when there is an assignee)</i></p> <p>10. <input type="checkbox"/> English Translation Document <i>(if applicable)</i></p> <p>11. <input type="checkbox"/> Information Disclosure Statement (IDS)/PTO-1449 <input type="checkbox"/> Copies of IDS Citations</p> <p>12. <input type="checkbox"/> Preliminary Amendment</p> <p>13. <input checked="" type="checkbox"/> Return Receipt Postcard (MPEP 503) <i>(Should be specifically itemized)</i></p> <p>14. <input type="checkbox"/> Small Entity <input type="checkbox"/> Statement filed in prior application, Status still proper and desired</p> <p>15. <input type="checkbox"/> Certified Copy of Priority Document(s) <i>(if foreign priority is claimed)</i></p> <p>16. <input checked="" type="checkbox"/> Other: <u>Bibliographic Data</u></p>	
<p>17. If a CONTINUING APPLICATION, check appropriate box and supply the requisite information:</p> <p><input type="checkbox"/> Continuation <input type="checkbox"/> Divisional <input type="checkbox"/> Continuation-in-part (CIP) of prior application No: _____</p>	

18. CORRESPONDENCE ADDRESS					
<input type="checkbox"/> Customer Number or Bar Code Label			<input checked="" type="checkbox"/> Correspondence address below		
(Insert Customer No. or Attach bar code label here)					
NAME	WESTERN DIGITAL CORPORATION				
	Milad G. Shara, Esq. - Reg. 39,367 <i>[Signature]</i> 6/19/00				
ADDRESS	8105 IRVINE CENTER DRIVE				
	PLAZA 3				
CITY	IRVINE	STATE	CALIFORNIA	ZIP CODE	92618
COUNTRY	U.S.A.	TELEPHONE	(949) 932-5676	FAX	(949) 932-5633

Burden Hour Statement: This form is estimated to take 0.2 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Box Patent Application, Washington, DC 20231.

Inventor Information

Inventor One Given Name :: Bruce E.
Family Name :: Aldridge
Name Suffix :: N/A
Postal Address Line One:: 1435 Calle Marbella
City :: Oceanside
State/Province :: CA
Country :: USA
Postal or Zip Code :: 92056
City of Residence :: Oceanside
Citizenship :: USA

Inventor Two Given Name :: Douglas J.
Family Name :: Durrant
Name Suffix :: N/A
Postal Address Line One:: 24061 Castilla Lane
City :: Mission Viejo
State/Province :: CA
Country :: USA
Postal or Zip Code :: 92691
City of Residence :: Mission Viejo
Citizenship :: USA

Correspondence Information

Name Line One :: Milad G. Shara, Esq.
Name Line Two :: Western Digital Corporation
Address Line One :: Plaza 3
Address Line Two :: 8105 Irvine Center Drive
City :: Irvine
State/Province :: California
Country :: USA
Postal or Zip Code :: 92618
Telephone :: (949) 932-5676
Fax :: (949) 932-5633
E-Mail :: Milad.G.Shara@wdc.com

Application Information

Title Line One :: Method And System For Aggregating Data Distribution Models
Formal Drawings :: Yes
Application Type :: Utility
Docket Number :: K35A0574
Licensed - U S Government Agency :: N/A
Contract Number :: N/A
Grant Number :: N/A
Secrecy Order in Parent Application :: N/A

Representative Information

Representative Customer Number One:: Milad G. Shara, Esq.
Registration Number One :: 39,367
Representative Customer Number Two:: Ivan M. Posey, Esq.
Registration Number Two:: 43,865

METHOD AND SYSTEM FOR AGGREGATING DATA DISTRIBUTION MODELS

BACKGROUND OF THE INVENTION

1. Field of the Invention

The invention relates to method and system for aggregating data distribution models. More specifically, the invention provides a system and method of aggregating or combining data distribution models while correctly maintaining the applicability of statistical and analytical techniques.

2. Description of the Prior Art and Related Information

When a researcher or engineer is involved in the analysis of a large quantity of data, called a data distribution, some summarization of the data elements in the data distribution is generally necessary because of the limitations of existing hardware and software. Generally this summarization of the distribution center and spread, including mean, sigma (standard deviation) and the number of data elements in the data distribution, are used. Alternatively sampling may be used to extract a smaller, more manageable subset of data that can be analyzed.

Unfortunately summarizing the data elements with a mean and sigma assumes that the distribution is Gaussian, or normal, and represents a single distribution and is not a mixture of independent distributions. Similarly, sampling may miss important elements of the distribution, for example outliers or bimodal patterns, unless the sample is sufficiently large.

Thus, there is a need for a system for advanced analysis that provides the benefit of maintaining the overall shape and characteristics of the data distribution while keeping the data storage requirements to a minimum. There is a further need for a system that can perform aggregation of small subgroups of the data distribution, thus keeping computation needs to a minimum. There is a further need for such a system with which statistical tests can be properly performed without making assumptions about the data distribution. There is a further need for a system with which complex analytics can be performed including basic statistical functions, such

as mean, minimum, maximum, standard deviation, etc. can be performed as well as complicated correlation and modeling studies. There is a further need for a system that naturally weights the highest data concentrations with the greatest accuracy in the approximation, wherein outliers are de-emphasized but not removed. There is a further need for a system with which an approximation of the original data distribution can be rebuilt from the model and estimates of the errors in this rebuilding can be made.

SUMMARY OF THE INVENTION

A system for creating an aggregated data model from a plurality data distribution models is disclosed. Each data distribution model is a summarized version of a data distribution having one or more data elements. Each data element has a value. Each data distribution model has one or more bins, wherein each bin approximates a subset of the data elements. Each bin comprises a start point having a value, an end point having a value, and a polynomial formula that approximates the data elements for the bin. Each data distribution model thus comprises a summarized representation of a data distribution, wherein the aggregated data model represents a combination of two or more of the data distribution models.

The system includes a processor for executing a computer program that is executable on a processor.

The computer program is adapted to perform a plurality of steps in a method for creating the aggregated data model. The computer program may contain a plurality of modules for performing the steps. One step comprises determining which start point has the minimum value and which end point has the maximum value of all of the bins of all of the data distribution models. The next step performed is setting a start point of a first bin of the aggregated data model to said start point determined to have the minimum value. The next step is setting an end point of a last bin of the aggregated data model to said end point determined to have the maximum value. The next step comprises determining a total number of points for the aggregated data model by adding the values indicating the number of data elements from all bins

1 from all data distribution models, each point comprising an approximated value of a data element
2 from one of the data distribution models. The next step comprises approximating the data
3 elements in the data distribution described by each data distribution model using the start point,
4 polynomial formula, and number of data elements for each bin in each respective data
5 distribution model, each approximated data element comprising one point in the aggregated data
6 model. The next step is to sort the points from minimum to maximum. The next step comprises
7 distributing the points into one or more bins in the aggregated data model such that a
8 substantially equal number of points are in each bin of the aggregated data model. The end point
9 of each bin in the aggregated data model may then be determined. The next step comprises
10 determining a polynomial formula for the points for each bin of the aggregated data model.

11 The computer program may create the bins of the aggregated data model on a bin by bin
12 basis. In that case the steps of approximating the data elements for the points of each bin,
13 determining the end point for each bin, and determining the polynomial formula for each bin are
14 performed for each bin individually.

15 **BRIEF DESCRIPTION OF THE DRAWINGS**

16 Fig. 1 is a block diagram illustrating the major components of a system for creating an
17 aggregated data model from a plurality data distribution models;

18 Fig. 2 is a flow diagram illustrating steps that may be performed by the system of Fig. 1;

19 Fig. 3 is a flow diagram illustrating steps that may be performed by the system of Fig. 1 for
20 creating each of the plurality of data distribution models that are aggregated by the system Fig. 1;

21 Fig. 4 is a flow diagram illustrating the steps performed by the computer program for
22 determining the start points and end points of the bins for each data distribution model according
23 to the method of Fig. 3 and system of Fig. 1;

24 Fig. 5 is a graphic illustration of two data distributions represented as a histogram

25 Fig. 6 is a graphic illustration of the data elements from one of the data distributions of Fig. 5
26 divided into bins;

Fig. 7 is a graphic illustration of an approximation of the original data distribution from Fig. 6 using the quadratic and spline fit verses a linear fit;
Fig. 8 is a graphic illustration showing the approximation error if a data distribution of Fig. 1 is treated as a normal distribution verses if the distribution is treated as a non-normal distribution using the system of Fig. 1; and
Fig. 9 is a graphic illustration of an aggregated data model aggregated from data distribution models of the data distributions of Fig. 5.

DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENTS

With reference to Fig. 1, a system for creating an aggregated data model 100 from a plurality data distribution models 102 is shown. Each data distribution model 80 is a summarized version of a data distribution 58 having one or more data elements 56, each data element 56 having a value, each data distribution model 102 having one or more bins 80 for approximating a subset of the data elements, each bin comprising a start point having a value, an end point having a value, and a polynomial formula approximating the data elements for the respective bin. Each data distribution model 102 thus comprises a summarized representation of a data distribution 58, wherein the aggregated data model 100 represents a combination of two or more of the data distribution models 102.

The system includes a processor 51 for executing a computer program 54 that is executable on a processor 51.

With reference to Fig. 2, the computer program 54 is adapted to perform a plurality of steps in a method for creating the aggregated data model 100. The computer program 54 may contain a plurality of modules 59 for performing the steps. One step comprises determining which start point has the minimum value and which end point has the maximum value of all of the bins 80 of all of the data distribution models 102, step 110. The next step performed is setting a start point of a first bin of the aggregated data model (a first bin of 180 in Fig. 180 described below) to said start point determined to have the minimum value, step 112. The next

1 step is setting an end point of a last bin 180 of the aggregated data model to said end point
2 determined to have the maximum value, step 114. The next step comprises determining a total
3 number of points for the aggregated data model by adding the values indicating the number of
4 data elements 56 from all bins from all data distribution models 102, each point comprising an
5 approximated value of a data element 56 from one of the data distribution models 102, step 116.
6 The next step comprises approximating the data elements 56 in the data distribution 85 described
7 by each data distribution model 102 using the start point, polynomial formula, and number of
8 data elements 56 for each bin 80 in each respective data distribution model 102, each
9 approximated data element 56 comprising one point in the aggregated data model 100, step 118.
10 The next step is to sort the points from minimum to maximum, step 120. The next step
11 comprises distributing the points into one or more bins (180 in Fig. 9) in the aggregated data
12 model 100 such that a substantially equal number of points are in each bin 180 of the aggregated
13 data model 100, step 122. The end point of each bin 180 in the aggregated data model 100 may
14 then be determined, step 124. The next step comprises determining a polynomial formula for the
15 points for each bin 180 of the aggregated data model 100, step 126.

16 The computer program may create the bins 180 of the aggregated data model 100 on a bin
17 by bin basis. In that case the steps of approximating the data elements 56 for the points of each
18 bin 180, determining the end point for each bin 180, and determining the polynomial formula for
19 each bin are performed for each bin 180 individually. The data elements 56 corresponding to the
20 bins 80 of the data distribution models 102 are thus approximated as needed using the respective
21 polynomial formula for the bin 80 of the respective data distribution model 102 in which the
22 needed data elements 56 are contained. This technique tends to conserve resources for the
23 processor 51 because once each bin 180 in the aggregated data model 100 is created, and the
24 polynomial formula for that bin 180 is determined, then the approximated data elements, or
25 points, for that particular bin 180 may be discarded before processing the next successive bin 180
26 for the aggregated data model.

1 The step of approximating the polynomial formula may comprise finding a quadratic
2 formula having the best fit with the points by using mathematical techniques such as the least
3 squares method. In the simplest case, the start and end points of the each bin 180 may be fit into
4 a linear formula by calculating the slope between the start and end point, and the y intercept
5 which is equal to the start point of the bin. Finding a polynomial formula is the preferred
6 method. The term polynomial formula as used herein may include a linear formula, quadratic
7 formula or other higher order polynomial formulas.

8 The step of distributing the points between, or into, the bins 180 in the aggregated data
9 model 100 may comprise dividing the number of total points in the aggregated data model 100
10 by the number of bins in the aggregated data model 100. If the number of points in the
11 aggregated data model 100 is not equally divisible by the number of bins 180, then the number of
12 points in each bin 180 is determined by dividing the number of points by the number of bins 180,
13 and then adding one to the count of the points in each of a number of bins 180 equal to the
14 remainder after dividing, wherein the bins 180 that have one added to the count is determined
15 according to the following formula:

16 for k from 1 to r

$$\text{bin}_{\text{add}} = \text{INT}((n * k) / (r + 1))$$

18 next k

19 wherein bin_{add} is the sequential bin number to add one to the count of points to include therein,
20 n is the total number of bins 180 in the aggregated data model 100, r is the remainder from
21 dividing the number of points in the aggregated data model 100 by the number of bins 180, and
22 INT is a function for rounding the result of the bracketed formula to produce an integer result.

23 Each of the data distribution models 102 that are used in aggregation may have been
24 created using one of several different methods. With reference to Fig. 3, a flow diagram
25 illustrating a method preformed by the computer program 54 for creating each of the one or more

1 data distribution models 102 from each of the one or more data distributions 58 is shown. The
2 data elements 56 are sorted from minimum to maximum, step 300. The number of data elements
3 56 in the data distribution 58 are computed, step 302. The value of the start point and the value
4 of the end point of each bin 80 are determined by distributing, or dividing, the data elements 56
5 into a plurality of substantially equal sized bins 80 for each data distribution 58, step 304, as
6 explained in more detail with respect to Fig. 4 below, step 304. The number of data elements 56
7 in each bin 80 are counted according to the following formula, step 306:

$$\text{start point} < \text{element value} \leq \text{end point}$$

8
9 wherein the start point is the start point of the respective bin 80, the element value is the value of
10 each data element 56 in each bin, and end point is the end point of the respective bin 80. The
11 data distribution model 102 may thus be computed by setting, for each bin 80, the start point of
12 the bin 80, the end point of the bin 80, and the number of data elements in the bin 80, step 308
13 for a linear model, or adding the step of determining a polynomial formula for a data distribution
14 model 102 that so uses one for approximating data elements 56 in each bin 80.

15 With reference to Fig. 4, a flow diagram illustrating the steps performed by the computer
16 program 54 for determining the start points and end points of the bins 80 for each data
17 distribution model 102 according to the method of Fig. 3 and system of Fig. 1 is shown. The start
18 point of the first bin 80 of the data distribution model 102 is selected as the value of the data
19 element 56 having the minimum value in the sorted data distribution 58, step 400. The start
20 point and end point of each bin 80 in the data distribution model 102 is determined according to
21 the following criteria, step 402:

- 22 (a) if the number of data elements 56 in the data distribution is equally divisible into the
23 number of bins 80, step 404, the end point of the first bin 80 is equal to the value of
24 the i th data element 56 in the data distribution 58, wherein i is the number of data
25 elements 56 in each bin determined by dividing the data elements 56 equally into the
26 number of bins 80, wherein the value of the end point of each bin 80 is equal to the

1 ith data element 56 after the last data element 56 in the proceeding bin 80, wherein
2 the start point of each bin 80 is equal to data element 56 after the last data element 56
3 of the previous bin 80, step 406, else

4 (b) if the number of data elements 56 in the data distribution 58 is not equally divisible
5 by the number of bins 80, then the number of data elements 56 in each bin 80 is
6 determined by dividing the number of data elements 56 by the number of bins 80, and
7 then adding one to the count of the data elements 56 in each of a number of bins 80
8 equal to the remainder after dividing, wherein the bins 80 that have one added to the
9 count is determined according to the following formula:

10 for k from 1 to r

11
$$\text{bin}_{\text{add}} = \text{INT}((n * k) / (r + 1))$$

12 next k

13 wherein bin_{add} is the sequential bin number to add one to the count of data elements to
14 include therein, n is the total number of bins 80 in the data distribution model, r is the remainder
15 from dividing the number of data elements 56 in the data distribution by the number of bins 80 in
16 the data distribution model, and INT is a function for rounding the result of the bracketed
17 formula to produce an integer result.

18 The computer program may perform the step of computing the number of data elements
19 56 in each bin 80 for the data distribution model 102 by counting, for each bin 80, each data
20 element 56 satisfying the following formula:

21
$$\text{start point} < \text{element value} \leq \text{end point}$$

22 wherein the bin start point is the start point of the respective bin 80, element value is the
23 value of each data element 56 in each bin 80, and end point is the end point of the respective bin
24 80.

1 A storage medium (70 in Fig. 1) may be provided for storing each data distribution model
2 102 by storing, for each bin 80, the start point, the end point, the number of data elements 56,
3 and the parameters of the polynomial formula that best approximates the data elements 56 for the
4 respective bin 80. Once the data distribution model 102 is stored, the original data distribution
5 58 from which the model was built no longer needs to be referred to. The computer program 54
6 may perform simple and complex statistical operations using the data model 102, or aggregations
7 of two or more data models 102. For example, the computer program may determine the range
8 of values of an aggregated data model 100 by subtracting the end point of the last bin in the
9 aggregated data model 100 from the start point of the first bin 180 in the aggregated data model
10 100, without having to refer to the original data elements 56 in the data distributions 58. The
11 computer program 54 may further determine the median value of the aggregated data model 100
12 by determining a number j computed by dividing the number of bins in the aggregated data
13 model by 2, and then reading the value of the end point of the j th bin as the median value if the
14 number of bins 180 in the aggregated data model 100 is equally divisible by 2 or by reading the
15 value of the mid point interpolated by the polynomial formula of the j th bin if the number of bins
16 in the aggregated data model 100 is not equally divisible by 2.

17 With reference to Fig. 5, an example of two data distributions 58 represented in
18 histograms representing the result of a measurement of stroke time, or time to sweep the heads
19 across the media, in magnetic or optical disk drives is shown. One data distribution 58 is for
20 product M, and one data distribution 58 is for product W. The bins 80 of the data distribution
21 models 102, shown as bars, give a histogram of the disk drives with a specific time interval
22 relative to the total population, shown as a fraction thereof, and each solid line 502 is a continuous
23 approximation of the data distribution 58 of the disk drives by time. In both cases integration of
24 the total area enclosed is equal to 1. Note that product M has more data elements 56 than product
25 W.

26 With reference to Fig. 6, a graph of the data elements 56 divided into bins 80 from the

data distribution 58 for product M from Fig. 5 is shown. The graph shows division of the curve into 10 bins (characterized by 11 end points). Information about the bins 80, including the number of data elements 56 in each bin 80 and parameters of the polynomial formula approximating the values of the data elements 56 within each bin 80 are stored in the storage medium 70 as shown in the table below.

Bin #	Start of Bin (c)	End of Bin	# of elements	Parameter 1 (a)	Parameter 2 (b)
1	8342 (min)		2701	3.82E-6	5.4E-8
2	8503		2705	1.76E-5	-6.1E-8
3	8540		2701	1.28E-5	-1.07E-7
4	8569		2674	6.5E-6	-1.26E-7
5	8596		2732	-4.00E-7	-1.23E-7
6	8624		2707	-7.50E-6	-9.88E-8
7	8655		2640	-1.39E-5	-4.95E-8
8	8693		2692	-1.25E-5	2.19E-8
9	9107		2693	-4.42E-7	4.69E-9
10	9542	10420 (max)	2694	-2.23E-6	1.87E-9

The table above shows the start points of the bins 80, which are the end points of preceding bins 80, the last end point of the last bin 80, and the parameters associated with a polynomial (spline) formula found using a spline fit for each bin 80 of the data elements 56 approximated from a data distribution 102. For example, the eighth record in the table indicates that the quadratic formula found by the computer program 54 to have the best fit comprises:

$$y = .0000125(x)^2 + .00000002.19(x) + 8693$$

In order to derive each polynomial formula, the computer program 54 may use techniques to fit the best approximation of the data elements 56 in each bin 80 of the respective data distribution 58 such as the least squares method, spline fit, linear fit, or other methods known to those skilled in the art for approximating the curve formed by the data elements 56.

The above table shows the minimum, maximum and median values of the data distribution 58 directly. Fig. 7 shows the approximation of the original data distribution 58 from Fig. 6 using the quadratic and spline fit verses a linear fit of the data distribution 58. As can be seen, a quadratic and spline fit is preferred because it offers a better approximation of the original raw data elements 56 that are also shown in Fig. 7. Typically it is expected that 100 bins 80 should be used for better fits. When 100 bins are used, linear interpolation between the end points of a bin 80 can be used – requiring less storage space in storage means 70 due to the higher bin density.

Thus, the data elements 56 from the original data distribution 58 can be reduced to $4*n+1$ points for cubic spline fits, $3*n+1$ points for quadratic fits, where n is the number of bins, or $2*n+1$ points for linear fits. Therefore the sample data can be reduced from, for example, 26,939 data elements to 31 for 10 bins using a quadratic fit to the data in each bin 80, or 201 elements using straight linear interpolation (linear fit) between the end points of the bins 80.

The minimum value, maximum value and range may be directly read from the distribution summary as the starting element 56 of the first bin 80, the ending element 56 of the last bin 80 and the difference between maximum and minimum.

The median is simply the value of the middle of the data distribution 58. It can be directly read by determining the value associated with the middle bin 80. For example the value of the end point of the 5th bin for 10 bins, or the 50th bin for 100 bins, or, for example, the interpolated middle of the 51st bin for 101 bins using the polynomial representation for the 51st bin.

The inter quartile range is important in various statistical analysis. It can be found by subtracting the value of the 25th percentile of the data which is the value of the end of the 25th bin for $n=100$ bins, from the value of the 75th percentiles, which is the value of the end of the 75th bin for $n=100$ bins.

1 Because the mean and standard deviation are technically only applicable for a Gaussian,
2 or normal, distributions, computation of these parameters may not be appropriate. If necessary
3 though the standard method for computation is to assume the distribution is normal. In that case
4 the median is equal to the mean and the standard deviation can be computed as the inter quartile
5 range divided by 1.349.

6 For determining outliers of the data distribution 58 the inter quartile range (IQR) is used.
7 Any data elements 56 greater than the value of the 75th percentile (i.e., the end point of bin 75)
8 plus 1.5* IQR can typically be considered an outlier.

9 With reference to Fig. 8, a graphic illustration showing the approximation error if a data
10 distribution 58 is treated as a normal distribution verses if the data distribution 58 is treated as a
11 non-normal distribution using the system of Fig. 1 is shown. Location 900 indicates a plot of the
12 standard deviation found with respect to the value of the data element 56 when plotting quantiles.
13 In contrast, location 903 shows same plot if a normal distribution is assumed. Other distributions
14 can be evaluated for error as well, for example, Weibull, lognormal, Poisson, F, Chi-square, etc.

15 With reference to Fig. 9, a data model 100 illustrating an aggregated data model 100
16 aggregated from the data distribution models 102 of the data distributions 58 of Fig. 5 is shown.
17 Linear interpolation and 100 bins are used to combine the data distribution models 102 from
18 product M and Product W as shown in Fig 1. Two tables were constructed to summarize the
19 distributions into 100 bins as shown in Appendix 1. The above described methods are then
20 applied to determine the start point, end point, and number of points for each bin 180.

Appendix 1: Summary tables for Product W, Product M and combined aggregation

Quantile	Product W			Product M			Combined		
	Bin Start	Bin Number	End	Bin Start	Bin Number	End	Bin Start	Bin Number	End
1%	7990	8121	158	8342	8427	271	7990	8155	426
2%	8121	8143	156	8427	8446	261	8155	8185	426
3%	8143	8160	156	8446	8459	275	8185	8204	425
4%	8160	8171	156	8459	8468	265	8204	8219	426
5%	8171	8181	156	8468	8475	241	8219	8232	425
6%	8181	8190	156	8475	8482	286	8232	8243	426
7%	8190	8197	156	8482	8488	285	8243	8254	425
8%	8197	8203	156	8488	8493	261	8254	8264	426
9%	8203	8209	156	8493	8498	269	8264	8272	425
10%	8209	8214	157	8498	8502	230	8272	8281	426
11%	8214	8219	156	8502	8507	304	8281	8289	425
12%	8219	8224	156	8507	8511	274	8289	8296	426
13%	8224	8229	156	8511	8515	261	8296	8303	425
14%	8229	8234	156	8515	8519	261	8303	8310	426
15%	8234	8238	156	8519	8523	282	8310	8317	425
16%	8238	8241	156	8523	8526	234	8317	8324	426
17%	8241	8246	156	8526	8529	247	8324	8331	425
18%	8246	8250	156	8529	8533	336	8331	8338	426
19%	8250	8254	157	8533	8536	230	8338	8345	425
20%	8254	8258	156	8536	8539	264	8345	8351	426
21%	8258	8261	156	8539	8542	246	8351	8358	425
22%	8261	8264	156	8542	8545	266	8358	8365	426
23%	8264	8267	156	8545	8548	267	8365	8371	425
24%	8267	8271	156	8548	8551	272	8371	8379	426
25%	8271	8274	156	8551	8554	273	8379	8386	425
26%	8274	8277	156	8554	8557	280	8386	8394	426
27%	8277	8280	156	8557	8560	321	8394	8402	425
28%	8280	8283	157	8560	8563	258	8402	8410	426
29%	8283	8286	156	8563	8566	281	8410	8419	425
30%	8286	8289	156	8566	8568	199	8419	8429	426
31%	8289	8292	156	8568	8571	297	8429	8437	425
32%	8292	8294	156	8571	8574	313	8437	8446	426
33%	8294	8297	156	8574	8576	201	8446	8453	425
34%	8297	8299	156	8576	8579	295	8453	8461	426
35%	8299	8302	156	8579	8582	280	8461	8468	425
36%	8302	8305	156	8582	8585	309	8468	8475	426
37%	8305	8307	157	8585	8587	193	8475	8482	425
38%	8307	8310	156	8587	8590	274	8482	8488	426

PATENT
Docket No. K35A0574

39%	8310	8313	156	8590	8593	316	8488	8494	425
40%	8313	8315	156	8593	8595	195	8494	8500	426
41%	8315	8317	156	8595	8598	298	8500	8506	425
42%	8317	8320	156	8598	8601	287	8506	8511	426
43%	8320	8323	156	8601	8604	265	8511	8517	425
44%	8323	8325	156	8604	8607	305	8517	8522	426
45%	8325	8328	156	8607	8609	214	8522	8527	425
46%	8328	8330	157	8609	8612	260	8527	8532	426
47%	8330	8333	156	8612	8615	313	8532	8537	425
48%	8333	8335	156	8615	8618	283	8537	8541	426
49%	8335	8338	156	8618	8620	204	8541	8546	425
50%	8338	8340	156	8620	8623	297	8546	8551	426
51%	8340	8343	156	8623	8626	305	8551	8555	425
52%	8343	8345	156	8626	8629	254	8555	8559	426
53%	8345	8348	156	8629	8632	289	8559	8563	425
54%	8348	8351	156	8632	8635	265	8563	8568	426
55%	8351	8353	157	8635	8638	260	8568	8572	425
56%	8353	8356	156	8638	8641	241	8572	8576	426
57%	8356	8358	156	8641	8645	342	8576	8580	425
58%	8358	8361	156	8645	8648	240	8580	8585	426
59%	8361	8363	156	8648	8650	300	8585	8589	425
60%	8363	8366	156	8650	8654	253	8589	8593	426
61%	8366	8368	156	8654	8657	222	8593	8598	425
62%	8368	8371	156	8657	8661	328	8598	8602	426
63%	8371	8374	156	8661	8664	221	8602	8606	425
64%	8374	8377	157	8664	8668	319	8606	8611	426
65%	8377	8380	156	8668	8671	232	8611	8615	425
66%	8380	8383	156	8671	8675	296	8615	8619	426
67%	8383	8386	156	8675	8679	252	8619	8623	425
68%	8386	8389	156	8679	8683	261	8623	8628	426
69%	8389	8391	156	8683	8687	272	8628	8632	425
70%	8391	8395	156	8687	8692	248	8632	8637	426
71%	8395	8398	156	8692	8697	294	8637	8642	425
72%	8398	8401	156	8697	8703	293	8642	8647	426
73%	8401	8404	157	8703	8709	247	8647	8651	425
74%	8404	8408	156	8709	8717	292	8651	8657	426
75%	8408	8411	156	8717	8724	254	8657	8662	425
76%	8411	8414	156	8724	8734	278	8662	8668	426
77%	8414	8418	156	8734	8748	281	8668	8673	425
78%	8418	8421	156	8748	8771	271	8673	8680	426
79%	8421	8425	156	8771	8848	274	8680	8686	425
80%	8425	8429	156	8848	9106	268	8686	8694	426
81%	8429	8433	156	9106	9338	270	8694	8702	425
82%	8433	8437	157	9338	9397	270	8702	8712	426

PATENT
Docket No. K35A0574

83%	8437	8441	156	9397	9426	269	8712	8724	425
84%	8441	8446	156	9426	9449	269	8724	8740	426
85%	8446	8450	156	9449	9467	270	8740	8770	425
86%	8450	8455	156	9467	9484	269	8770	8931	426
87%	8455	8460	156	9484	9499	270	8931	9235	425
88%	8460	8465	156	9499	9514	269	9235	9397	426
89%	8465	8471	156	9514	9528	269	9397	9439	425
90%	8471	8478	156	9528	9541	270	9439	9468	426
91%	8478	8484	157	9541	9554	269	9468	9493	425
92%	8484	8492	156	9554	9567	269	9493	9517	426
93%	8492	8503	156	9567	9580	270	9517	9538	425
94%	8503	8515	156	9580	9595	269	9538	9558	426
95%	8515	8530	156	9595	9611	270	9558	9579	425
96%	8530	8562	156	9611	9632	269	9579	9602	426
97%	8562	8689	156	9632	9654	269	9602	9632	426
98%	8689	9158	156	9654	9684	270	9632	9670	425
99%	9158	9819	156	9684	9728	269	9670	9761	425
100%	9819	11018	156	9728	10420	269	9761	11018	426

CLAIMS

WHAT IS CLAIMED IS:

1. A system for creating an aggregated data model from a plurality data distribution models, each data distribution model describing a data distribution having one or more data elements, each data element having a value, each data distribution model having one or more bins, each bin comprising a start point having a value, an end point having a value, a value indicating the number of data elements for each bin, and a polynomial formula associated with each bin, the polynomial formula approximating the data elements for the respective bin, comprising:
 - a processor; and
 - a computer program executable on a processor, the computer program adapted to perform the following steps:
 - (a) determining which start point has the minimum value and which end point has the maximum value of all of the bins of all of the data distribution models;
 - (b) setting a start point of a first bin of the aggregated data model to said start point determined to have the minimum value;
 - (c) setting an end point of a last bin of the aggregated data model to said end point determined to have the maximum value;
 - (d) determining a total number of a plurality of points for the aggregated data model by adding the values indicating the number of data elements from all bins from all data distribution models;
 - (e) approximating the data elements in the data distribution described by each data distribution model using the start point, polynomial formula, and number of data elements for each bin in each respective data distribution

model, each approximated data element comprising one of said points in the aggregated data model;

- (f) sorting the points from minimum to maximum;
- (g) distributing the points into one or more bins in the aggregated data model such that a substantially equal number of points are in each bin of the aggregated data model; and
- (h) determining a polynomial formula with the sorted data elements for each bin of the aggregated data model.

2. The system of claim 1, wherein the computer program is further for determining the end point for each bin in the aggregated data model.

3. The system of claim 1, wherein the computer program is adapted to perform the step of distributing the points into the one or more bins of the aggregated data model according to the following formula::

- (a) if the number of points in the aggregated data model is equally divisible into the number of bins, the end point of the first bin is equal to the value of the i th point in the aggregated data model, wherein i is the number of points in each bin determined by dividing the points equally into the number of bins, wherein the value of the end point of each bin is equal i th point after the last point in the proceeding bin, wherein the start point of each bin is equal to the point after the last point of the previous bin, else
- (b) if the number of data elements in the points is not equally divisible by the number of bins, then the number of points in each bin is determined by dividing the number of points by the number of bins, and then adding one to the count of the points in each of a number of bins equal to the

remainder after dividing, wherein the bins that have one added to the count is determined according to the following formula:

for k from 1 to r

$$\text{bin}_{\text{add}} = \text{INT}((n * k) / (r + 1))$$

next k

wherein bin_{add} is the sequential bin number to add one to the count of points to include therein, n is the total number of bins in the aggregated data model, r is the remainder from dividing the number of points in the data distribution by the number of bins, and INT is a function for rounding the result of the bracketed formula to produce an integer result.

4. The system of claim 1, wherein the computer program is for performing separately for each bin of the aggregated data model, the steps of approximating the data elements for each bin, determining the end point for each bin, and determining the polynomial formula for each bin.

5. The system of claim 1, wherein each data distribution model is the result of the computer program performing a the following steps:

- (a) sorting the data elements in from minimum to maximum for each data distribution;
- (b) computing the number of data elements in each data distribution;
- (c) determining the value of the start point and the value of the end point of each bin by dividing the data elements into a plurality of substantially equal sized bins for each data distribution;
- (d) counting the number of data elements in each bin for each data distribution; and

- 11 (e) computing each distribution model for each data distribution, each
12 distribution model comprising, for each bin, the start point of the bin, the
13 end point of the bin, and the number of data elements in the bin.

1 6. The system of claim 5, wherein the computer program is adapted to perform the
2 following steps for determining the start points and end points of the bins for each data
3 distribution model:

4 selecting as the start point of the first bin the value of the data element having the
5 minimum value in the sorted data distribution;

6 determining the start point and end point of each bin according to the following
7 criteria:

- 8 (c) if the number of data elements in the data distribution is equally divisible
9 into the number of bins, the end point of the first bin is equal to the value
10 of the *i*th data element in the data distribution, wherein *i* is the number of
11 data elements in each bin determined by dividing the data elements
12 equally into the number of bins, wherein the value of the end point of each
13 bin is equal *i*th data element after the last data element in the proceeding
14 bin, wherein the start point of each bin is equal to the data element after
15 the last data element of the previous bin, else

- 16 (d) if the number of data elements in the data distribution is not equally
17 divisible by the number of bins, then the number of data elements in each
18 bin is determined by dividing the number of data elements by the number
19 of bins, and then adding one to the count of the data elements in each of a
20 number of bins equal to the remainder after dividing, wherein the bins that
21 have one added to the count is determined according to the following
22 formula:

for k from 1 to r

$\text{bin}_{\text{add}} = \text{INT}((n * k) / (r + 1))$

next k

wherein bin_{add} is the sequential bin number to add one to the count of data elements to include therein, n is the total number of bins in the data distribution model, r is the remainder from dividing the number of data elements in the data distribution by the number of bins, and INT is a function for rounding the result of the bracketed formula to produce an integer result.

7. The system of claim 6, wherein the computer program is further for performing the step of counting by counting, for each bin, each data element satisfying the following formula:

$\text{start point} < \text{element value} \leq \text{end point}$

wherein the bin start point is the start point of the respective bin, element value is the value of each data element in each bin, and end point is the end point of the respective bin.

8. The system of claim 7, comprising a storage medium for storing each data distribution model by storing, for each bin, the start point, the end point, the number of data elements, and the parameters of the polynomial formula.

9. The system of claim 1, wherein the computer program is further for performing one or more statistical analysis using the aggregated data model.

10. The system of claim 9, wherein the statistical analysis performed comprises determining the range of the points of the aggregated data model analyzed by subtracting end point of

the last bin in the aggregated data model from the start point of the first bin in the aggregated data model.

11. The system of claim 9, wherein the statistical analysis performed comprises determining the inter quantile range of the points of the aggregated data model.

12. The system of claim 9, wherein the statistical analysis performed comprises determining the median value of the aggregated data model by determining a number j computed by dividing the number of bins by 2, and then reading the value of the end point of the j th bin as the median value if the number of bins in the aggregated data model is equally divisible by 2 or by reading the interpolated value using the polynomial function of the mid point of the j th bin if the number of bins in the aggregated data model is not equally divisible by 2.

13. A method for creating an aggregated data model from a plurality data distribution models, each data distribution model describing a data distribution having one or more data elements, each data element having a value, each data distribution model having one or more bins, each bin comprising a start point having a value, an end point having a value, a value indicating the number of data elements for each bin, and a polynomial formula associated with each bin, the polynomial formula approximating the data elements for the respective bin, the method comprising:

determining which start point has the minimum value and which end point has the maximum value of all of the bins of all of the data distribution models;

setting a start point of a first bin of the aggregated data model to said start point determined to have the minimum value;

setting an end point of a last bin of the aggregated data model to said end point determined to have the maximum value;

determining a total number of a plurality of points for the aggregated data model by adding the values indicating the number of data elements from all bins from all data distribution models;

approximating the data elements in the data distribution described by each data distribution model using the start point, polynomial formula, and number of data elements for each bin in each respective data distribution model, each approximated data element comprising one of said points in the aggregated data model;

sorting the points from minimum to maximum;

distributing the points into one or more bins in the aggregated data model such that a substantially equal number of points are in each bin of the aggregated data model; and

determining a polynomial formula with the sorted data elements for each bin of
the aggregated data model.

14. The method of claim 13, comprising determining the end point for each bin in the
aggregated data model.

15. The method of claim 13, wherein the step of distributing the points into the one or more
bins of the aggregated data model is performed according to the following formula::

- (e) if the number of points in the aggregated data model is equally divisible
into the number of bins, the end point of the first bin is equal to the value
of the i th point in the aggregated data model, wherein i is the number of
points in each bin determined by dividing the points equally into the
number of bins, wherein the value of the end point of each bin is equal i th
point after the last point in the proceeding bin, wherein the start point of
each bin is equal to the point after the last point of the previous bin, else
(f) if the number of data elements in the points is not equally divisible by the
number of bins, then the number of points in each bin is determined by
dividing the number of points by the number of bins, and then adding one
to the count of the points in each of a number of bins equal to the
remainder after dividing, wherein the bins that have one added to the count
is determined according to the following formula:

for k from 1 to r

$$\text{bin}_{\text{add}} = \text{INT}(n * k) / (r + 1))$$

next k

wherein bin_{add} is the sequential bin number to add one to the count of
points to include therein, n is the total number of bins in the aggregated

data model, and r is the remainder from dividing the number of points in the data distribution by the number of bins, and INT is a function for rounding the result of the bracketed formula to produce an integer result.

16. The method of claim 13, wherein the steps of approximating the data elements for each bin, determining the end point for each bin, and determining the polynomial formula for each bin are performed separately for each bin of the aggregated data model.

17. The method of claim 13, comprising creating each data distribution model using the following steps:

- (f) sorting the data elements in from minimum to maximum for each data distribution;
- (g) computing the number of data elements in each data distribution;
- (h) determining the value of the start point and the value of the end point of each bin by dividing the data elements into a plurality of substantially equal sized bins for each data distribution;
- (i) counting the number of data elements in each bin for each data distribution; and
- (j) computing each distribution model for each data distribution, each distribution model comprising, for each bin, the start point of the bin, the end point of the bin, and the number of data elements in the bin.

18. The method of claim 17, comprising determining the start points and end points of the bins for each data distribution model using the following steps:
selecting as the start point of the first bin the value of the data element having the minimum value in the sorted data distribution;

determining the start point and end point of each bin according to the following criteria:

(g) if the number of data elements in the data distribution is equally divisible into the number of bins, the end point of the first bin is equal to the value of the i th data element in the data distribution, wherein i is the number of data elements in each bin determined by dividing the data elements equally into the number of bins, wherein the value of the end point of each bin is equal i th data element after the last data element in the proceeding bin, wherein the start point of each bin is equal to the data element after the last data element of the previous bin, else

(h) if the number of data elements in the data distribution is not equally divisible by the number of bins, then the number of data elements in each bin is determined by dividing the number of data elements by the number of bins, and then adding one to the count of the data elements in each of a number of bins equal to the remainder after dividing, wherein the bins that have one added to the count is determined according to the following formula:

for k from 1 to r

$$\text{bin}_{\text{add}} = \text{INT}(n * k) / (r + 1))$$

next k

wherein bin_{add} is the sequential bin number to add one to the count of data elements to include therein, n is the total number of bins in the data distribution model, r is the remainder from dividing the number of data elements in the data distribution by the number of bins, and INT is a

function for rounding the result of the bracketed formula to produce an integer result.

19. The method of claim 18, wherein the step of counting is performed by counting, for each bin, each data element satisfying the following formula:

$$\text{start point} < \text{element value} \leq \text{end point}$$

wherein the bin start point is the start point of the respective bin, element value is the value of each data element in each bin, and end point is the end point of the respective bin.

20. The method of claim 19, comprising storing each data distribution model by storing, for each bin, the start point, the end point, the number of data elements, and the parameters of the polynomial formula.

21. The method of claim 13, comprising performing one or more statistical analysis using the aggregated data model.

22. The method of claim 21, wherein the statistical analysis performed comprises determining the range of the points of the aggregated data model analyzed by subtracting end point of the last bin in the aggregated data model from the start point of the first bin in the aggregated data model.

23. The method of claim 21, wherein the statistical analysis performed comprises determining the inter quantile range of the points of the aggregated data model.

24. The method of claim 21, wherein the statistical analysis performed comprises determining the median value of the aggregated data model by determining a number j computed by dividing the number of bins by 2, and then reading the value of the end

- 4 point of the jth bin as the median value if the number of bins in the aggregated data
5 model is equally divisible by 2 or by reading the interpolated value using the polynomial
6 function of the mid point of the jth bin if the number of bins in the aggregated data model
7 is not equally divisible by 2.

METHOD AND SYSTEM FOR AGGREGATING DATA DISTRIBUTION MODELS

ABSTRACT

A system and method for creating an aggregated data model from a plurality data distribution models having bins approximating data elements in a plurality of data distributions is disclosed. Each bin of each data distribution model has a polynomial formula for approximating data elements in a respective data distribution. A method for creating the aggregated data model comprises: determining a start point of the aggregated data model having the minimum value and an end point having a maximum value of all of the bins of all of the data distribution models; setting a start point of a first bin of the aggregated data model; setting an end point of a last bin of the aggregated data model; determining a total number of points for the aggregated data model; approximating the data elements described by each data distribution model, each approximated data element comprising one point in the aggregated data model; sorting the points from minimum to maximum; distributing the points into one or more bins in the aggregated data model; and determining a polynomial formula for the points for each bin of the aggregated data model.

Fig. 2

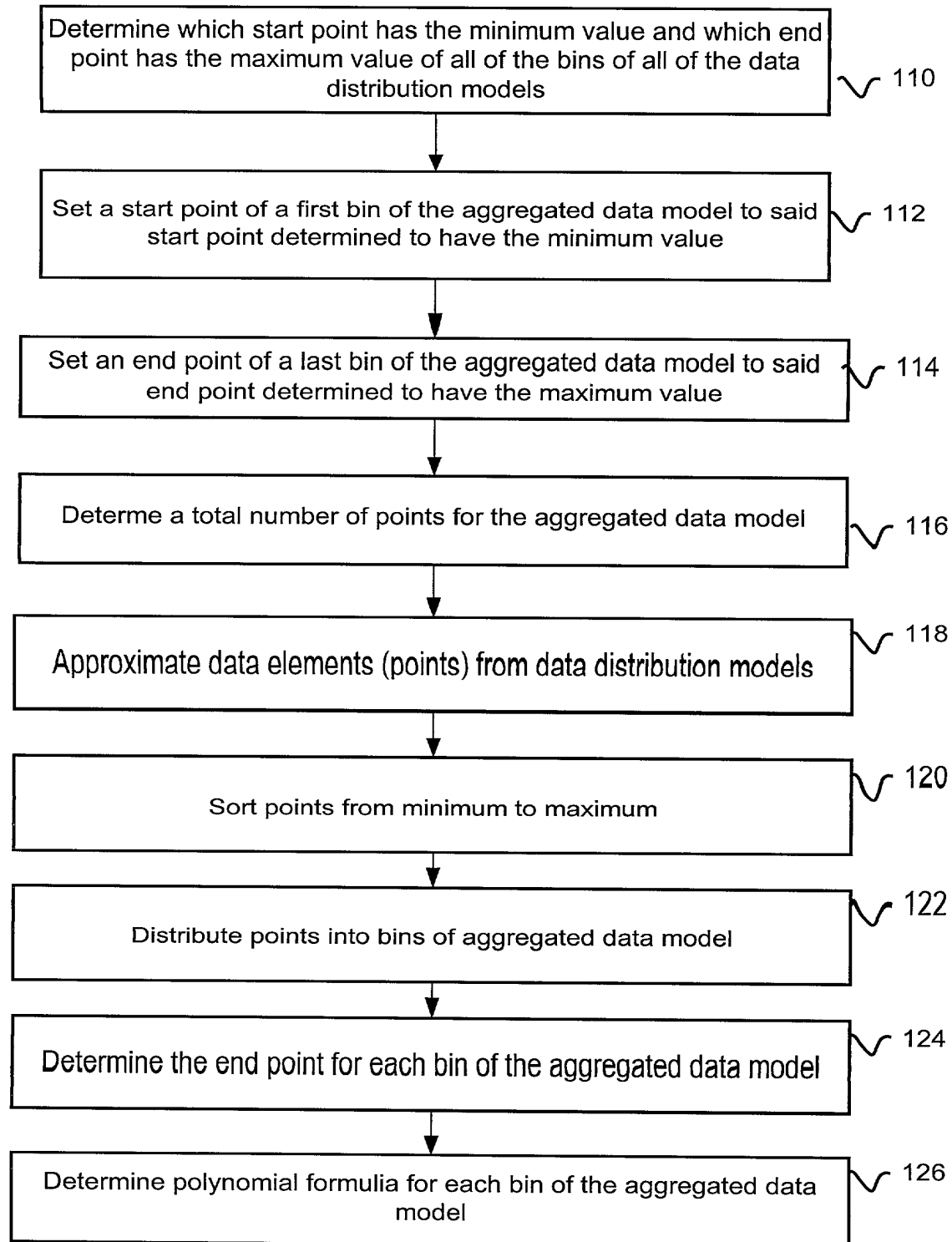


Fig. 3

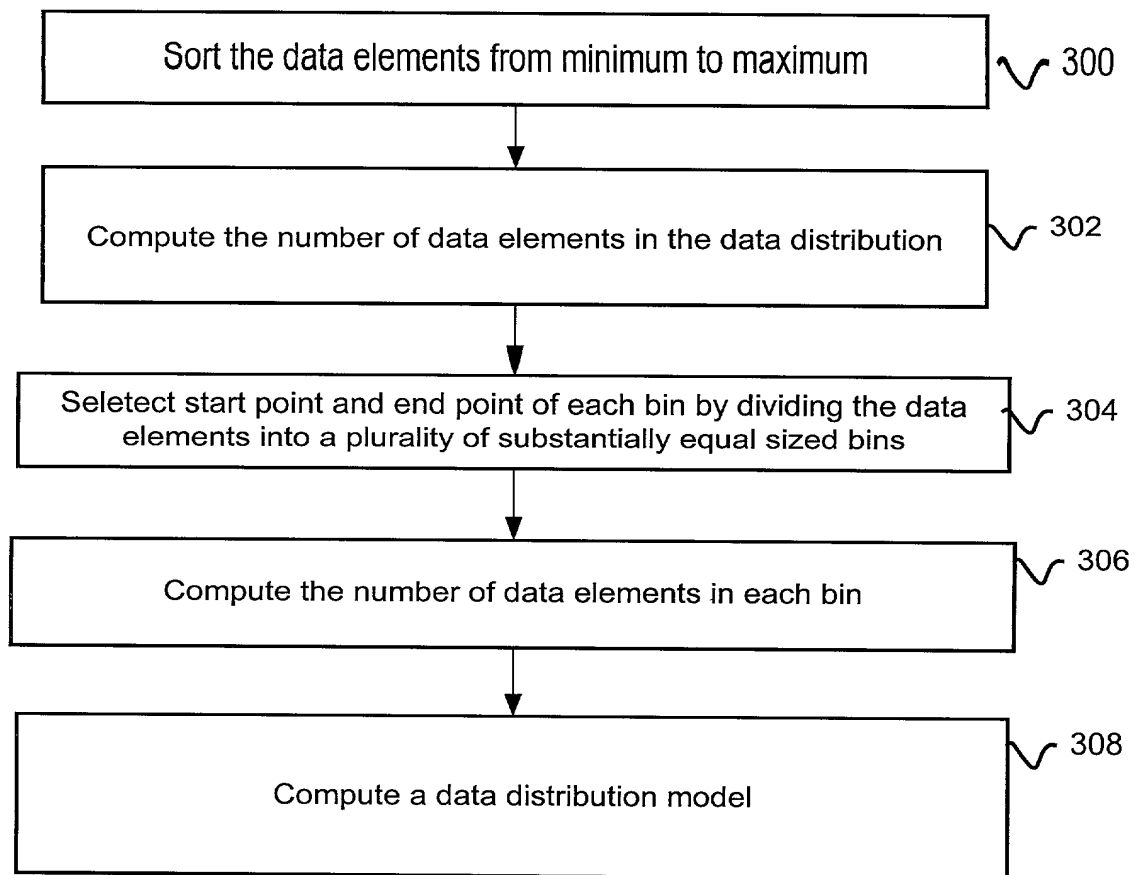


Fig. 4

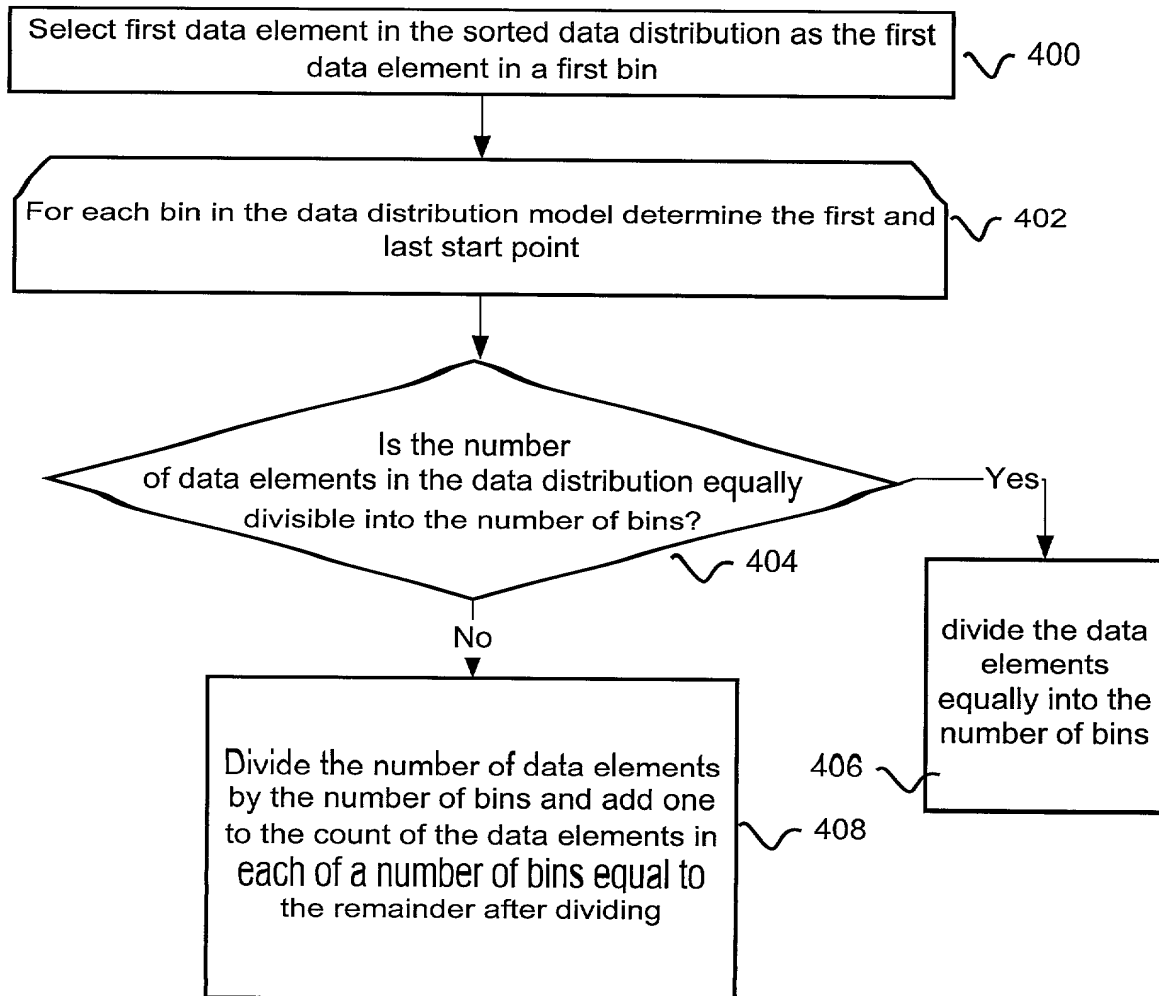


Fig. 5

Stroke time - Product M vs. Product W

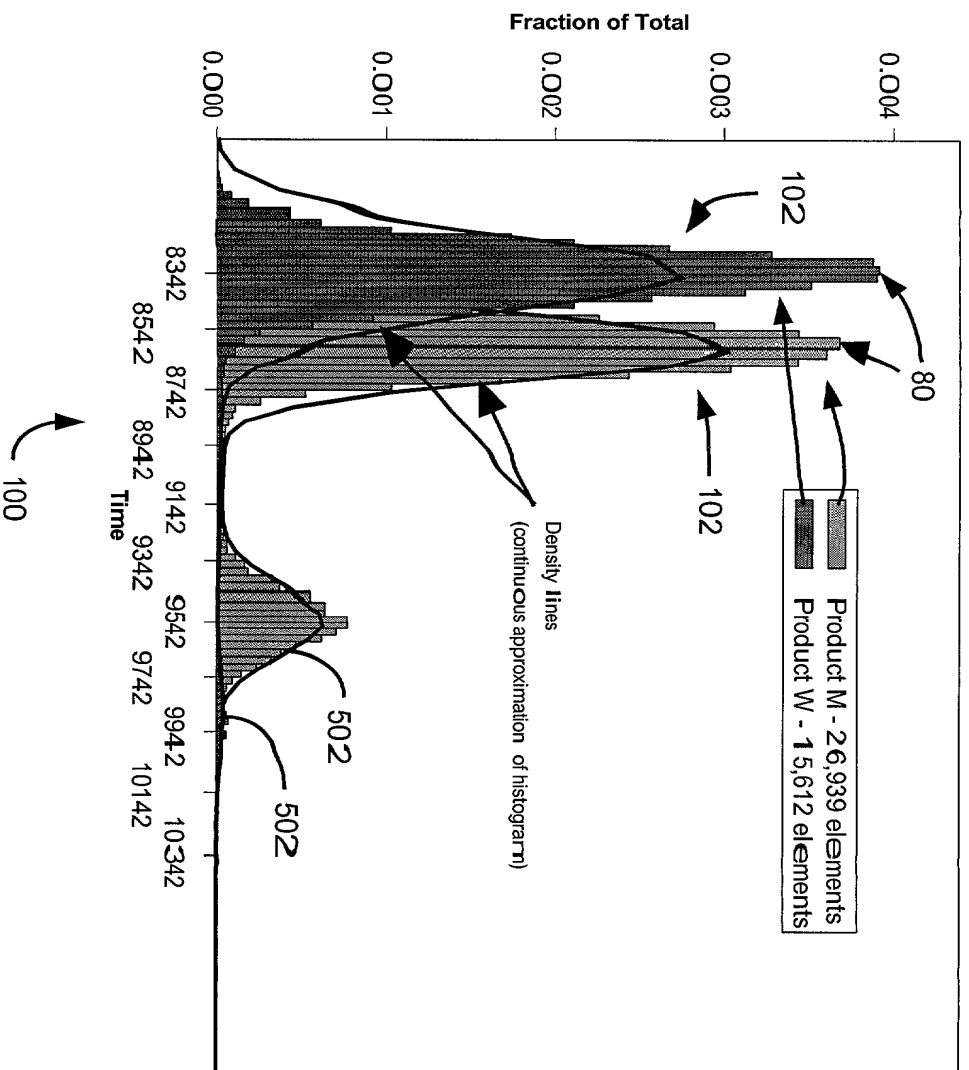
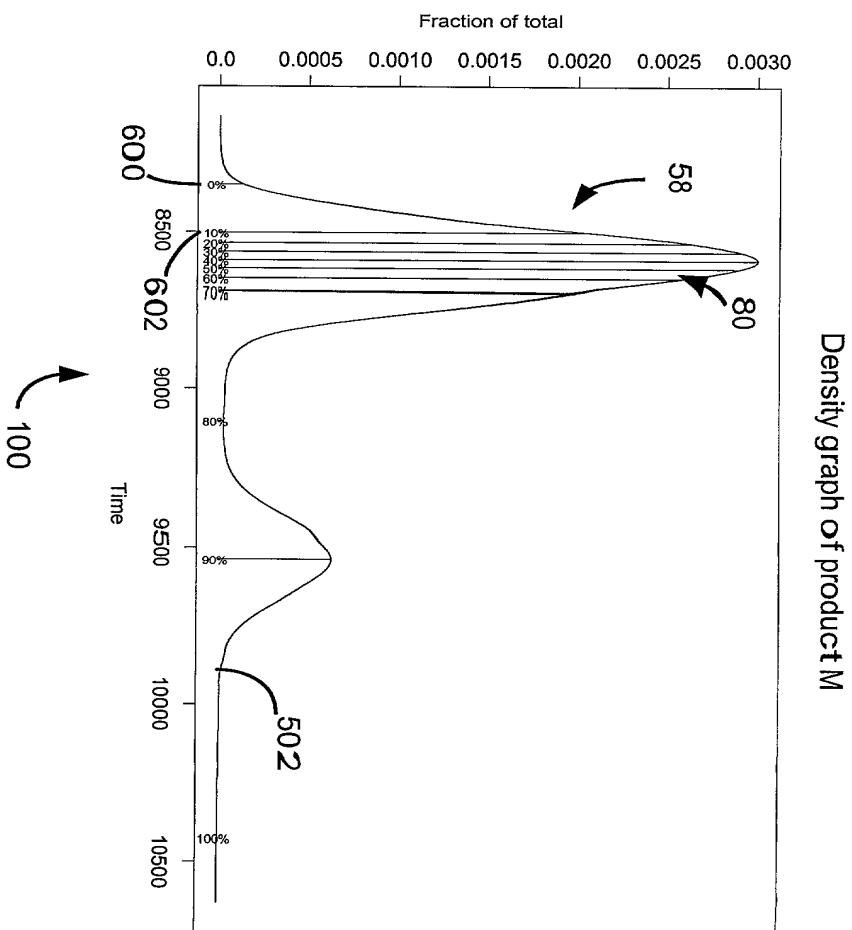


Fig. 6



Data Count vs. linear and quadratic quantile fits

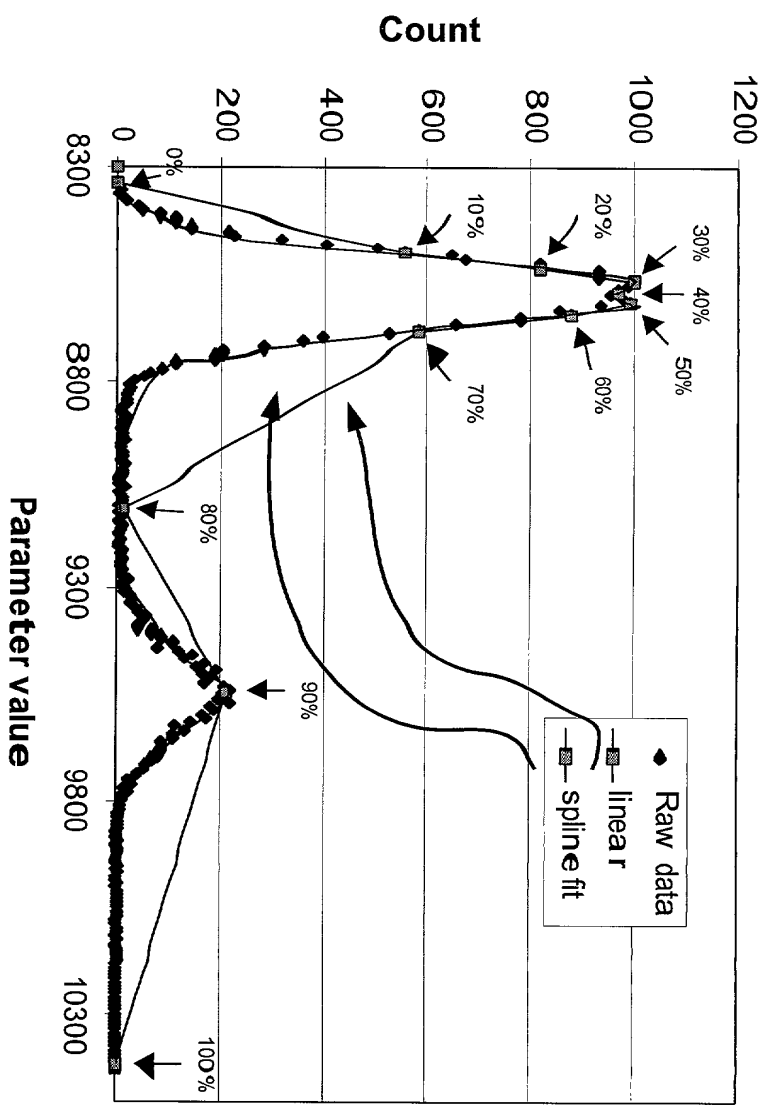


Fig. 7

Fig. 8

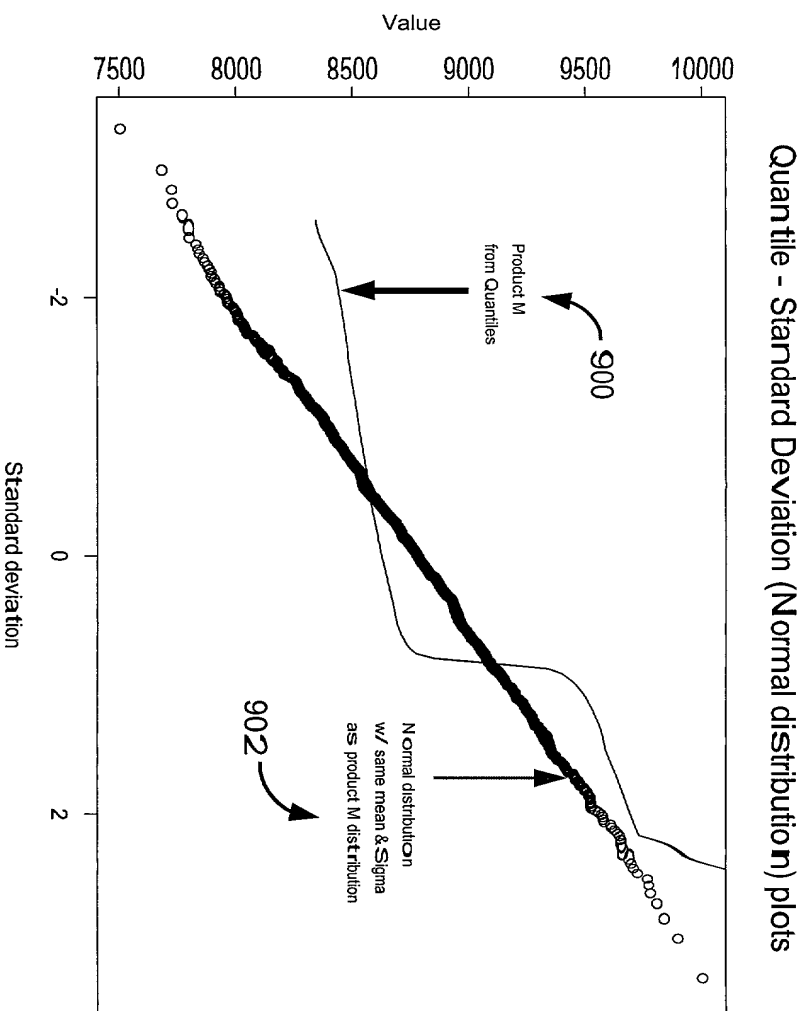
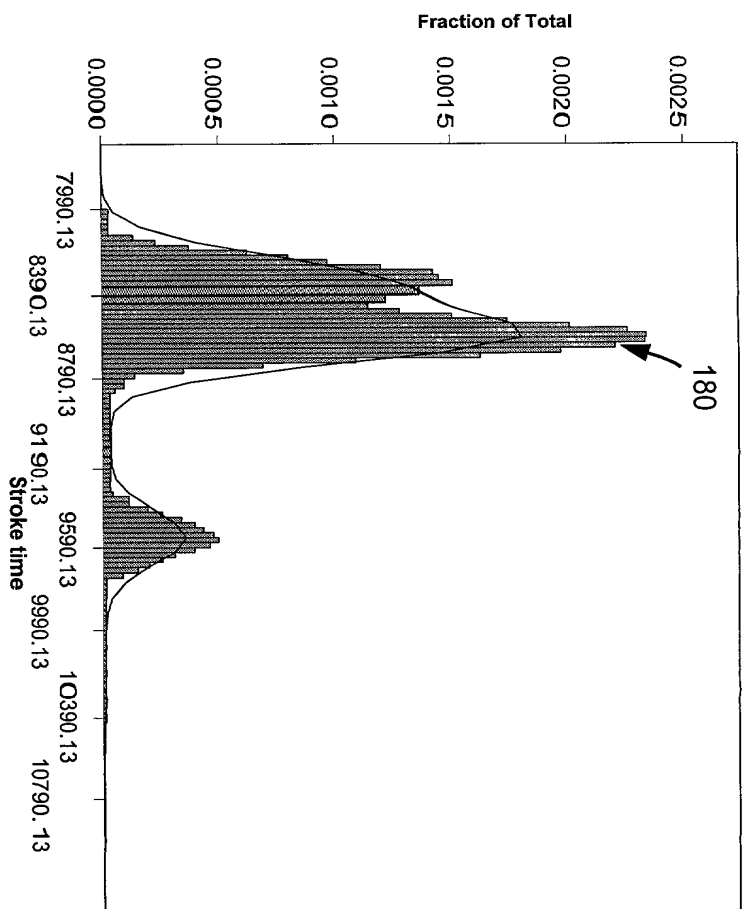


Fig. 9

Combined distribution of Product W and Product M



09590.13 061000

Please type a plus sign (+) inside this box → ☐

PTO/SB/01 (12-97)
Approved for use through 9/30/00. OMB 0651-0032
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

DECLARATION FOR UTILITY OR DESIGN PATENT APPLICATION (37 CFR 1.63) <input checked="" type="checkbox"/> Declaration Submitted with Initial Filing OR <input type="checkbox"/> Declaration Submitted after Initial Filing (surcharge (37 CFR 1.16 (e)) required)	Attorney Docket Number	K35A0574
	First Named Inventor	BRUCE E. ALDRIDGE
	COMPLETE IF KNOWN	
	Application Number	/ Unknown
	Filing Date	Herewith
	Group Art Unit	Unknown
	Examiner Name	Unknown

As a below named inventor, I hereby declare that:

My residence, post office address, and citizenship are as stated below next to my name.

I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural names are listed below) of the subject matter which is claimed and for which a patent is sought on the invention entitled:

METHOD AND SYSTEM FOR AGGREGATING DATA DISTRIBUTION MODELS

the specification of which (Title of the Invention)

☒ is attached hereto
OR

☐ was filed on (MM/DD/YYYY) as United States Application Number or PCT International

Application Number and was amended on (MM/DD/YYYY) (if applicable).

I hereby state that I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment specifically referred to above.

I acknowledge the duty to disclose information which is material to patentability as defined in 37 CFR 1.56.

I hereby claim foreign priority benefits under 35 U.S.C. 119(a)-(d) or 365(b) of any foreign application(s) for patent or inventor's certificate, or 365(a) of any PCT international application which designated at least one country other than the United States of America, listed below and have also identified below, by checking the box, any foreign application for patent or inventor's certificate, or of any PCT international application having a filing date before that of the application on which priority is claimed.

Prior Foreign Application Number(s)	Country	Foreign Filing Date (MM/DD/YYYY)	Priority Not Claimed	Certified Copy Attached?	
				YES	NO
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
			<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

☐ Additional foreign application numbers are listed on a supplemental priority data sheet PTO/SB/02B attached hereto:

I hereby claim the benefit under 35 U.S.C. 119(e) of any United States provisional application(s) listed below.

Application Number(s)	Filing Date (MM/DD/YYYY)	<input type="checkbox"/> Additional provisional application numbers are listed on a supplemental priority data sheet PTO/SB/02B attached hereto.

[Page 1 of 2]

Burden Hour Statement: This form is estimated to take 0.4 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

Please type a plus sign (+) inside this box →



PTO/SB/01 (12-97)

Approved for use through 9/30/00. OMB 0651-0032

Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

DECLARATION — Utility or Design Patent Application

I hereby claim the benefit under 35 U.S.C. 120 of any United States application(s), or 365(c) of any PCT international application designating the United States of America, listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States or PCT International application in the manner provided by the first paragraph of 35 U.S.C. 112, I acknowledge the duty to disclose information which is material to patentability as defined in 37 CFR 1.56 which became available between the filing date of the prior application and the national or PCT international filing date of this application.

U.S. Parent Application or PCT Parent Number	Parent Filing Date (MM/DD/YYYY)	Parent Patent Number (if applicable)

☐ Additional U.S. or PCT international application numbers are listed on a supplemental priority data sheet PTO/SB/02B attached hereto.

As a named inventor, I hereby appoint the following registered practitioner(s) to prosecute this application and to transact all business in the Patent and Trademark Office connected therewith:

☐ Customer Number

OR

☒ Registered practitioner(s) name/registration number listed below

Place Customer
Number Bar Code
Label here

Name	Registration Number	Name	Registration Number
Milad G. Shara	39,367		
Ivan M. Posey	43,865		

☐ Additional registered practitioner(s) named on supplemental Registered Practitioner Information sheet PTO/SB/02C attached hereto.

Direct all correspondence to: ☐ Customer Number or Bar Code Label ☒ Correspondence address below

Name	Milad G. Shara				
Address	WESTERN DIGITAL CORPORATION				
Address	8105 Irvine Center Drive, Plaza 3				
City	Irvine	State	California	ZIP	92618
Country	U.S.A.	Telephone	(949) 932-5676	Fax	(949) 932-5633

I hereby declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under 18 U.S.C. 1001 and that such willful false statements may jeopardize the validity of the application or any patent issued thereon.

Name of Sole or First Inventor: ☐ A petition has been filed for this unsigned inventor

Given Name (first and middle [if any])		Family Name or Surname			
BRUCE E.		ALDRIDGE			
Inventor's Signature	<i>Bruce E. Aldridge</i>			Date	6/16/00
Residence: City	OCEANSIDE	State	CA	Country	USA
Post Office Address	1435 CALLE MARBELLA				
Post Office Address					
City	Oceanside	State	CA	ZIP	92056
				Country	USA

☒ Additional inventors are being named on the 1 supplemental Additional Inventor(s) sheet(s) PTO/SB/02A attached hereto

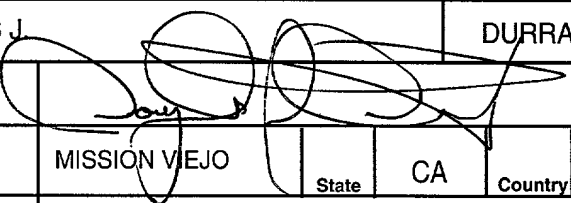
Please type a plus sign (+) inside this box → +

PTO/SB/02A (3-97)
Approved for use through 9/30/98. OMB 0651-0032
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE
Under the Paperwork Reduction Act of 1995, no persons are required to respond to a collection of information unless it contains a valid OMB control number.

+

DECLARATION

ADDITIONAL INVENTOR(S)
Supplemental Sheet
Page 1 of 1

Name of Additional Joint Inventor, if any:				<input type="checkbox"/> A petition has been filed for this unsigned inventor			
Given Name (first and middle [if any])				Family Name or Surname			
DOUGLAS J.				DURRANT			
Inventor's Signature				Date		6.16.00	
Residence: City		MISSION VIEJO		State		CA	
				Country		USA	
Post Office Address		24061 CASTILLA LANE					
Post Office Address							
City		MISSION VIEJO		State		CA	
				ZIP		92691	
				Country		USA	
Name of Additional Joint Inventor, if any:				<input type="checkbox"/> A petition has been filed for this unsigned inventor			
Given Name (first and middle [if any])				Family Name or Surname			
Inventor's Signature				Date			
Residence: City				State		CA	
				Country		USA	
Post Office Address							
Post Office Address							
City				State		CA	
				ZIP			
				Country		USA	
Name of Additional Joint Inventor, if any:				<input type="checkbox"/> A petition has been filed for this unsigned inventor			
Given Name (first and middle [if any])				Family Name or Surname			
Inventor's Signature				Date			
Residence: City				State		CA	
				Country		USA	
Post Office Address							
Post Office Address							
City				State		CA	
				ZIP			
				Country		USA	

Burden Hour Statement: This form is estimated to take 0.4 hours to complete. Time will vary depending upon the needs of the individual case. Any comments on the amount of time you are required to complete this form should be sent to the Chief Information Officer, Patent and Trademark Office, Washington, DC 20231. DO NOT SEND FEES OR COMPLETED FORMS TO THIS ADDRESS. SEND TO: Assistant Commissioner for Patents, Washington, DC 20231.

+